A BI-DIRECTIONAL DUAL PROMOTER COMPLEX WITH ENHANCED PROMOTER ACTIVITY FOR TRANSGENE EXPRESSION IN EUKARYOTES

The present application is a non-provisional application claiming priority under 35 USC 119(e) to U.S. Provisional Application No. 60/268,358, of Li et al., entitiled A BI-DIRECTIONAL DUAL PROMOTER COMPLEX WITH ENHANCED PROMOTER ACTIVITY FOR TRANSGENE EXPRESSION IN EUKARYOTES, filed February 13, 2001, which is incorporated herein in its entirety by reference.

The present invention relates to bidirectional dual promoter complexes (BDPC)for enhancement of transgene expression. More particularly, a BDPC is constructed by placing two core promoters on either side of modified enhancers.

BACKGROUND

Gene expression is composed of several major processes, including transcription, translation and protein processing. Among these processes, transcription not only dictates the precise copying of DNA into mRNA but also provides sophisticated mechanisms for the control of gene expression. There are a number of fundamental steps involved in transcription: promoter recognition and binding by transcription factors and RNA polymerase components, nascent RNA chain initiation, RNA transcript elongation, and RNA transcript termination (Uptain et al., Ann. Rev. Biochem. 66:117-172 (1997)). Promoters are an essential component for transcription, effecting transcription both quantitatively and qualitatively. A promoter contains numerous DNA motifs or cis-elements that can serve as recognition signals and binding sites for transcription factors. Working

together with transcription factors, these cis-elements can function as architectural elements or anchoring points for achieving promoter geometry (Perez-Martin et al., Ann. Rev. Microbiol. 51:593-628 (1997)).

Numerous promoters have been isolated from a wide variety of organisms ranging from viruses to animals. They have become the subjects of intensive studies in efforts to characterize their molecular organization and the basic mechanisms regulating transcriptional control of gene expression. In recent years, a number of well-characterized promoters have been successfully adopted for use in the genetic transformation of plants. These promoters control transgene expression in transgenic plants and have been used in efforts to improve agronomic performance and to incorporate value-added features. However, in spite of the availability of these promoters, there is currently a shortage of promoters for use in genetic transformation research with plants. In most instances, use of existing plant promoters isolated from a specific species to effect transformation in a different species results in reduced promoter activity and/or altered patterns of gene expression, reflecting the variation of genetic background between different species (Ellis et al., EMBO J. 6:11-16 (1987); Miao et al., Plant Cell 3:11-22 (1991)). Recently, a constitutive actin gene promoter isolated from *Arabidopsis* (An et al., Plant J. 10:107-121 (1996)) failed to support desired levels of transgene expression in grape cells. To date, the promoter most commonly used to effect transformation in crop plants is the cauliflower mosaic virus 35S (CaMV 35S) promoter and its derivatives (Sanfacon, Can. J. Bot. 70:885-899 (1992)). The CaMV 35S promoter was originally isolated from a plant virus.

Successful genetic transformation of plants frequently requires the use of more than one promoter to

adequately drive expression of multiple transgenes. For instance, at least three promoters are normally needed in order to express a selectable marker gene, a reporter marker gene and a target gene of interest. Multiple promoters are required because almost all the mRNAs in eukaryotes are monocistronic (single polypeptide-encoding transcript). Hence, expression of complex traits controlled by more than a single target gene in plants has been thought to require the use of additional promoters.

Recent studies have showed that foreign DNA integrated into the plant genome can be recognized by host factors and that the foreign DNA may be subsequently subjected to modifications that lead to transgene silencing. Mechanisms involved in this process include; DNA methylation, chromatin structural modification and post-transcriptional mRNA degradation (Kumpatla et al., TIBS 3:97-104 (1998)). In general, foreign DNA containing repeated sequences, including sequences homologous to host DNA, is more prone to gene silencing modifications (Selker, Cell 97:157-160 (1999)). Accordingly, the repeated use of the same promoter in transformation vector may increase the probability of gene silencing and unstable transgene expression in transgenic plants. As more transgenic crop plants are developed for release to the farmers, transgene silencing is likely to become a major concern. Hence, there is an urgent need to develop new promoters that will efficiently drive transgene expression, especially in transgenic plants.

Over the years, several strategies have been adopted for use to improve the performance of various promoters. These strategies can be classified into two categories, namely 1) modification of homologous promoters and 2) construction of heterologous promoters.

Modification of homologous promoters is accomplished by manipulating the enhancer region of a particular promoter in an effort to achieve higher transcriptional activity without altering existing expression patterns. Kay et al. (Science 236:1299-1302 (1987) first demonstrated that approximately ten-fold higher transcriptional activity was achieved by tandem duplication of 250 base pairs of the upstream enhancer region of the CaMV 35S promoter, as compared to the transcriptional activity of the natural promoter. Mitsuhara et al. (Plant Cell Physiol. 37:49-59 (1996)) further showed that other forms of tandem repeats of the upstream enhancer region of the CaMV 35S promoter were also capable of producing 10 to 50 fold higher levels of transgene expression in rice and tobacco without altering the constitutive expression pattern of the promoter.

Modification of promoters using heterologous enhancer sequences is also commonly practiced to achieve higher transcriptional activity and desired expression patterns. For example, a CaMV 35S promoter upstream enhancer fragment was fused to the nopaline synthase promoter (NOS) and the resulting fusion promoter reportedly increased the transcriptional activity, as compared to the weaker NOS promoter (Odell, et al. PMB 10:263-272 (1988)). The upstream enhancer regions of the CaMV 35S promoter and the octopine synthase promoter were used to fuse with the maize Adhl promoter to enhance transcription activity, while retaining the anaerobic regulation pattern of the Adhl promoter (Ellis et al. EMBO J.6:11-16 (1987) and 6:3203-3208 (1987)). The achievement of transcriptional enhancement by using heterologous enhancers is primarily attributable to the unique characteristics of enhancers, which could exert its functions to regulate transcriptional activity in an orientation- and position-independent fashion.

SUMMARY

The present invention is directed to a bidirectional dual promoter complex (BDPC) for enhancement of transgene expression and a method for constructing a BDPC. In accordance with the invention, the BDPC includes at least two core promoters and at least one modified internal enhancer region. The core promoters are fused to either end of the modified enhancer region in a divergent orientation such that the transcriptional direction (5' to 3') of each promoter points away from each other (see for example Fig. 1). The modified enhancer region includes at least two tandem oriented enhancer sequences having substantial sequence identity. Each core promoter is capable of independently directing transcription of a transgene that may contain expressible or nonexpressible coding sequences.

In another aspect of the invention, both enhancer and core promoter components used in a BDPC may be derived from homologous and/or heterologous promoter sequences. More specifically, in a homologous BDPC, the repeated enhancer sequences and core promoters may be isolated from a single source promoter that is composed of an enhancer and a core promoter. In a heterologous BDPC, the repeated enhancer sequences may be isolated from a promoter source that is different from that which the source promoter from which the core promoters are obtained.

The core promoter of the present invention includes a DNA sequence that corresponds to about 50 bp to about 100 bp. The core promoter may include a TATA-box consensus element and an Initiator (INR). In another aspect of the invention, the core promoter includes a TATA-box consensus element, an INR, and at least one cis-acting element such as a CAAT-box or an as-1 element (Benfey et al., Science 250:959-966 (1990)). Core promoters in a BDPC may have substantial sequence

identity or in one aspect of the invention, be identical. In another aspect, the core promoters of the invention may have a sequence homology of at least about 30% and include at least 5 bp identical, contiguous nucleotides within the core promoter region.

The modified enhancer region in the BDPC may include at least two enhancer sequences having substantial sequence identity arranged in a tandem orientation. In one aspect, the enhancer sequences are identical. The modified enhancer regions are constructed such that the 3' end of a first enhancer sequence is linked to the 5' end of a second enhancer sequence to form a modified enhancer region of the BDPC of the invention. In another aspect, more than two, or multiples of two, such as four and six, repeated enhancer sequences can also be used to construct a BDPC. In an aspect of the invention where four enhancer sequences are used, a first tandem two-unit enhancer region may be fused with another tandem two-unit enhancer region in a back-to-back orientation. The DNA sequence of each enhancer region in a BDPC may be about 100 bp to about 1.0 kbp. In one aspect, transcriptional efficiency is increased when enhancer regions are asymmetrical. The size of an enhancer region is based on desired requirements for the level of transcriptional activity and on desired requirements for a specific transgene expression regulation mechanism.

The modified enhancer region of the BDPC of the invention may also include enhancer sequences that are fully functional to the core promoters used in the BDPC. In this aspect of the invention, enhancers that are fully functional are capable of modulating, including enhancing or down regulating, the initiation and synthesis of transcripts from a transgene containing either translatable or non-translatable coding sequences.

In another aspect, the BDPC of the invention is utilized to provide simultaneous control of transgene

transcription and expression from both core promoters whose transcriptional activities are significantly enhanced by the arrangement of the promoter complex. The use of the BDPC of the invention in transgenic hosts is effective for providing enhanced levels of transcription in both transient expression and stable transformation assays. In this aspect of the invention, by using a homologous BDPC that includes two modified enhancer regions and two core promoters, all of which are derived from the same source promoter, up to a 220-fold increase in transcriptional activity was obtained from an upstream core promoter as compared to transcriptional activity from the same core promoter alone (see Fig. 13). Up to a 2-fold increase in transcription activity can be achieved from an upstream core promoter in a BDPC as compared to that same core promoter having the same enhancer sequences but not in a BDPC. Further, transcriptional activity may be increased as much as 40% in a downstream core promoter in a BDPC as compared to a double enhancer with a core promoter.

In another aspect, the present invention is effective for increasing the number of transcription units and for enhancing transcription control based on the use of a limited number of promoter sequences. Since DNA sequences from a single promoter source can be used to construct a homologous BDPC for the expression of two, or more than two in the case of translation fusion, monocistronic transgene sequences, the number of promoters required to express multiple transgenes is reduced by using the BDPC of the invention. In addition, expression of these multiple transgenes is under the control of the same BDPC and regulated simultaneously according to regulatory information encoded within the shared enhancer region and core promoters. Accordingly, the BDPC of the present invention is effective for achieving synchronized expression of complex multi-gene-

controlled quantitative traits loci (QTL), including those responsible for major events of growth and development in crop plants and other higher organisms. In this aspect, the invention provides transgenic plants, asexual cuttings from these plants in certain instances, and seeds from transgenic plants in certain instances, that contain the BDPC of the present invention. The BDPC of the present invention are also effective for reducing transcriptional silencing of transgene expression.

Examples of BDPCs are set forth in Figure 2 (SEQ. ID. Nos.: 1 and 2), Figure 4 (SEQ. ID. Nos.: 3 and 4), Figure 6 (SEQ. ID. Nos.: 5 and 6), Figure 8 (SEQ. ID. No.: 7 and 8), Figure 10 (SEQ. ID. No.: 9 and 10) Figure 12 (SEQ. ID. No.: 11 and 12), Figure 19 (SEQ. ID. No.: 13 and 14), Figure 21 (SEQ. ID. No.: 15 and 16), and Figure 23 (SEQ. ID. No.: 17 and 18).

BRIEF DESCRIPTION OF FIGURES

Figure 1 illustrates a BDPC with 2 enhancers based on CaMV 35S promoter.

Figure 2 shows the nucleotide sequence (SEQ. ID. Nos.: 1 and 2) of the BDPC illustrated in Figure 1.

Figure 3 illustrates a BDPC with 4 enhancers based on CaMV 35S promoter.

Figure 4 shows the nucleotide sequence (SEQ. ID. Nos.: 3 and 4) of the BDPC illustrated in Figure 3.

Figure 5 illustrates a BDPC with 2 enhancers based on CsVMV promoter.

Figure 6 shows the nucleotide sequence (SEQ. ID. Nos.: 5 and 6) of the BDPC illustrated in Figure 5.

Figure 7 illustrates a BDPC with 4 enhancers based on CsVMV promoter.

Figure 8 shows the nucleotide sequence (SEQ. ID. Nos.: 7 and 8) of the BDPC illustrated in Figure 7.

Figure 9 illustrates a BDPC with 2 enhancers based on ACT2 promoter.

Figure 10 shows the nucleotide sequence (SEQ. ID. Nos.: 9 and 10) of the BDPC illustrated in Figure 9.

Figure 11 illustrates a BDPC with 2 enhancers based on PRb1b promoter of tobacco.

Figure 12 shows the nucleotide sequence (SEQ. ID. Nos.: 11 and 12) of the BDPC illustrated in Figure 11.

Figure 13 illustrates a physical map of the T-DNA region of binary vectors containing a BDPC.

Figure 14 illustrates transient GFP expression in grape SE (somatice embryo, *Vitis vinifera* cv. Thompson Seedless) after transformation using binary vectors p201 and p201R.

Figure 15 shows transient GFP expression efficiency of grape SE (*Vitis vinifera* cv. Thompson Seedless) after transformation using binary vectors p201 and p201R.

Figure 16 shows an analysis of GUS activity in grape SE (*Vitis vinifera* cv. Thompson Seedless) after transformation using binary vectors p201 and p201R.

Figure 17 illustrates GFP expression in grape SE(A) and leaf tissue (B) of transgenic grape (*Vitis vinifera* cv. Thompson Seedless) containing the T-DNA of p201R.

Figure 18 illustrates a BDPC with 2 enhancers based on At UBQ1 promoter.

Figure 19 shows the nucleotide sequence (SEQ. ID. Nos.: 13 and 14) of the BDPC illustrated in Figure 18.

Figure 20 illustrates a heterologous BDPC with 2 UBQ-1 enhancers and 2 CsVMV core promoters.

Figure 21 shows the nucleotide sequence (SEQ. ID. Nos.: 15 and 16) of the BDPC illustrated in Figure 20.

Figure 22 illustrates a heterologous BDPC with 2 PR1b enhancers and 2 CaMV 35S core promoters.

Figure 23 shows the nucleotide sequence (SEQ. ID. Nos.: 17 and 18) of the BDPC illustrated in Figure 22.

Figure 24 illustrates a physical map of a T-DNA region of CaMV 35S promoter-derived binary vectors containing a BDPC.

Figure 25 shows the analysis of GUS activity in three different grape SE (*V. Vinifera cv.* Thompson Seedless) lines after transformation using three binary vectors.

Figure 26 illustrates a physical map of a T-DNA region of transformation vectors with 4-enhancer-containing BDPC.

Figure 27 shows the analysis of GUS activity in SE (*V. Vinifera cv.* Thompson Seedless) lines after transformation using three binary vectors.

DETAILED DESCRIPTION

Definitions

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Singleton et al. (1994) Dictionary of Microbiology and Molecular Biology, second edition, John Wiley and Sons (New York) provides one of skill with a general dictionary of many of the terms used in this invention. All patents and publications referred to herein are incorporated by reference herein. For purposes of the present invention, the following terms are defined below.

The term "nucleic acid" refers to a deoxyribonucleotide or ribonucleotide polymer in either single- or double-stranded form, or sense or anti-sense, and unless otherwise limited, encompasses known analogues of natural nucleotides that hybridize to nucleic acids in manner similar to naturally occurring nucleotides. Unless otherwise indicated, a particular nucleic acid sequence includes the complementary sequence thereof.

The terms "operably linked", "in operable combination", and "in operable order" refer to functional linkage between a nucleic acid expression control sequence (such as a promoter, signal sequence, or array

of transcription factor binding sites) and a second nucleic acid sequence, wherein the expression control sequence affects transcription and/or translation of the nucleic acid corresponding to the second sequence. In the present application, the gene of interest that is operably linked to the BDPC may be upstream or downstream from the BDPC.

The term "recombinant" when used with reference to a cell indicates that the cell replicates a heterologous nucleic acid, expresses said nucleic acid or expresses a peptide, heterologous peptide, or protein encoded by a heterologous nucleic acid. Recombinant cells can express genes that are not found within the native (non-recombinant) form of the cell. Recombinant cells can also express genes that are found in the native form of the cell, but wherein the genes are modified and re-introduced into the cell by artificial means.

A "structural gene" is that portion of a gene comprising a DNA segment encoding a protein, polypeptide or a portion thereof, and excluding the 5' sequence which drives the initiation of transcription. The structural gene may alternatively encode a nontranslatable product. The structural gene may be one which is normally found in the cell or one which is not normally found in the cell or cellular location wherein it is introduced, in which case it is termed a "heterologous gene". A heterologous gene may be derived in whole or in part from any source known to the art, including a bacterial genome or episome, eukaryotic, nuclear or plasmid DNA, cDNA, viral DNA or chemically synthesized DNA. A structural gene may contain one or more modifications which could effect biological activity or the characteristics, the biological activity or the chemical structure of the expression product, the rate of expression or the manner of expression control. Such modifications include, but are not limited to, mutations, insertions, deletions and

substitutions of one or more nucleotides. The structural gene may constitute an uninterrupted coding sequence or it may include one or more introns, bounded by the appropriate splice junctions. The structural gene may be translatable or non-translatable, including in an anti-sense orientation. The structural gene may be a composite of segments derived from a plurality of sources (naturally occurring or synthetic, where synthetic refers to DNA that is chemically synthesized).

"Divergent orientation" refers to an arrangement where sequences are pointing away from each other or in opposite directions in their direction of transcription.

"Derived from" is used to mean taken, obtained, received, traced, replicated or descended from a source (chemical and/or biological). A derivative may be produced by chemical or biological manipulation (including, but not limited to, substitution, addition, insertion, deletion, extraction, isolation, mutation and replication) of the original source.

"Chemically synthesized", as related to a sequence of DNA, means that the component nucleotides were assembled in vitro. Manual chemical synthesis of DNA may be accomplished using well established procedures (Caruthers, Methodology of DNA and RNA Sequencing, (1983), Weissman (ed.), Praeger Publishers, New York, Chapter 1); automated chemical synthesis can be performed using one of a number of commercially available machines.

Two polynucleotides or polypeptides are said to be "identical" if the sequence of nucleotides or amino acid residues in the two sequences is the same when aligned for maximum correspondence. Optimal alignment of sequences for comparison may be conducted by the local homology algorithm of Smith and Waterman, Adv. Appl. Math. 2:482 (1981), by the homology alignment algorithm of Needleman and Wunsch, J. Mol. Biol. 48:443 (1970), by the search for similarity method of Pearson and Lipman

Proc. Natl. Acad. Sci. (U.S.A.) 85: 2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, Wis.), or by inspection.

The terms "substantial identity" or "substantial sequence identity" as applied to nucleic acid sequences and as used herein denote a characteristic of a polynucleotide sequence, wherein the polynucleotide comprises a sequence that has at least 85 percent sequence identity, preferably at least 90 to 95 percent sequence identity, and more preferably at least 99 percent sequence identity as compared to a reference sequence over a comparison window of at least 20 nucleotide positions, frequently over a window of at least 25-50 nucleotides, wherein the percentage of sequence identity is calculated by comparing the reference sequence to the polynucleotide sequence which may include deletions or additions which total 20 percent or less of the reference sequence over the window of comparison. The reference sequence may be a subset of a larger sequence.

Another indication that nucleotide sequences are substantially identical is if two molecules hybridize to each other under stringent conditions. Stringent conditions are sequence-dependent and will be different in different circumstances. Generally, stringent conditions are selected to be about 5°C to about 20°C, usually about 10°C to about 15°C, lower than the thermal melting point (Tm) for the specific sequence at a defined ionic strength and pH. The Tm is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. Typically, stringent conditions will be those in which the salt concentration is about 0.02 molar at pH 7 and the temperature is at least about 60°C. For instance in

a standard Southern hybridization procedure, stringent conditions will include an initial wash in 6xSSC at 42 °C followed by one or more additional washes in 0.2xSSC at a temperature of at least about 55°C, typically about 60°C and often about 65°C.

Nucleotide sequences are also substantially identical for purposes of this invention when the polypeptides which they encode are substantially identical. Thus, where one nucleic acid sequence encodes essentially the same polypeptide as a second nucleic acid sequence, the two nucleic acid sequences are substantially identical, even if they would not hybridize under stringent conditions due to silent substitutions permitted by the genetic code (see, Darnell et al. (1990) Molecular Cell Biology, Second Edition Scientific American Books W. H. Freeman and Company New York for an explanation of codon degeneracy and the genetic code).

Protein purity or homogeneity can be indicated by a number of means well known in the art, such as polyacrylamide gel electrophoresis of a protein sample, followed by visualization upon staining. For certain purposes high resolution will be needed and HPLC or a similar means for purification utilized.

As used herein, the term "cis" is used in reference to the presence of nucleic acid signal binding elements on a chromosome. The term "cis-acting" is used in reference to the controlling effect of a regulatory nucleic acid element on a gene. For example, enhancers and promoters may include cis acting control elements which may affect transcription.

As used herein, the term "vector" is used in reference to nucleic acid molecules that transfer DNA segment(s) into a cell. A vector may act to replicate DNA and may reproduce independently in a host cell. The term "vehicle" is sometimes used interchangeably with "vector."

The term "expression vector" as used herein refers to a recombinant DNA molecule containing a desired coding sequence and appropriate nucleic acid sequences necessary for the expression of the operably linked coding sequence in a particular host organism. Nucleic acid sequences necessary for expression in prokaryotes usually include a promoter, an operator (optional), and a ribosome binding site, often along with other sequences. Eucaryotic cells are known to utilize promoters, enhancers, and termination and polyadenylation signals.

As used herein, the term "TATA element" or "TATA box" is used in reference to a segment of DNA, located approximately 19-27 base pairs upstream from the transcription start point of eucaryotic structural genes, to which RNA polymerase binds. The TATA box is approximately 7 base pairs in length, often comprising as one example, the sequence "TATAAAA" or "TATATAA". The TATA box is also sometimes referred to as the "Hogness box."

The term "CAAT box" or "CAAT element" refers to a conserved DNA sequence located upstream from the TATA box or the transcription start point of eucaryotic structural genes, to which RNA polymerase binds.

Transcriptional control signals in eukaryotes comprise "promoter" and "enhancer" elements. Promoters and enhancers consist of short arrays of DNA sequences that interact specifically with cellular proteins involved in transcription (Maniatis, T. et al., Science 236:1237 (1987)). Promoter and enhancer elements have been isolated from a variety of eukaryotic sources including genes in yeast, insect and mammalian cells, plants and viruses (analogous control elements, i.e., promoters, are also found in prokaryotes). The selection of a particular promoter and enhancer depends on what cell type is to be used to express the protein of interest. Some eukaryotic promoters and enhancers have a

broad host range while others are functional in a limited subset of cell types (for review see Voss, S. D. et al., Trends Biochem. Sci., 11:287 (1986) and Maniatis, T. et al., supra (1987)).

As used herein the term "transgene" refers to any gene that is not normally present in a particular host.

"Expressible coding sequence", as used herein, refers to a DNA sequence that serves as a template for the synthesis gene products or polypeptides. "Non-expressible coding sequence" refers to any DNA sequences that direct the synthesis of non-translatable transcripts, including antisense mRNA.

Core Promoters

In an important aspect, the BCPC of the present invention includes at least two core promoters. Structurally, the term "core promoter", as used herein, may correspond to, but not limited to, a DNA sequence of about 50 bp to about 100 bp in length. The DNA sequence may contain at least a TATA-box consensus element and the Initiator (INR), and preferably a TATA-box consensus element, the INR and at least one cis-acting element such as the CAAT-box or the *as-1* element (Benfey and Chua, Science 250:959-966 (1990)). A core promoter may be commonly isolated from DNA sequences immediately upstream of a transcription start site (TSS) or synthesized chemically according to pre-determined DNA sequence information.

Functionally, the term "core promoter", as used herein, is defined by its capability to direct the precise initiation and synthesis of transcripts from an operably linked nucleic acid sequence at a minimum activity level that can be detected by using currently available gene transcription analysis methods, including reverse transcriptase-polymerase chain reaction assay (RT-PCR), nucleic acid hybridization techniques, DNA-protein binding assays and in vitro and/or in vivo gene

expression analysis approaches using living cells (Wefald, et al., Nature 344:260-262 (1990); Benfey and Chua, Science 250:959-966 (1990); Patikoglou and Burley, annu. Rev. Biophys. Biomol. Struct. 26:289-325 (1997)). In one aspect, the core promoters of the invention have a sequence homology where promoter sequences have a homology when compared to each other of at least about 30% and include at least 5 bp identical contiguous nucleotides within the core promoter region.

Both structural and functional features of various core promoters have been previously studied extensively and described in great details in literature (Kollmar and Farnham, Proc. Exp. Biol. Med. 203:127-139 (1993); Orphanides, et al. Genes and Dev. 10:2657-2683 (1996); Roeder, Trends Biochem. Sci. 21:327-335 (1996); Tjian, Philos. Trans. R. Soc. Lond. B. Biol. Sci. 351:491-499 (1996)).

A core promoter is generally referred to as a DNA sequence that is directly located upstream of a nucleic acid sequence that is to be transcribed. However, in a BDPC said nucleic acid sequence may be either upstream or downstream from a core promoter. The nucleic acid sequence to be transcribed may be either translatable or non-translatable and may further include an open reading frame or coding sequence.

The TATA-box and the INR are the two key elements present in a core promoter, both of which play an important role in determining the TSS position and in initiating basal transcription. The consensus sequence for the TATA-box may comprise TATA(A/T)A(A/T) and the INR has the consensus YYAN(T/A)YY, where the underlined A indicates the TSS. According to observations from numerous cloned gene promoters, abundantly expressed genes generally contain a strong TATA-box in their core promoter, while most housekeeping genes, including oncogenes and those encoding growth factors and

transcription factors, may often contain no TATA-box in their core promoter. In some strong core promoters, other cis-acting elements, including the CAAT-box and the *as-1* element, are frequently found to be overlapped within the core promoter DNA sequence. For instance, the core promoter of the CaMV 35S promoter was defined experimentally to be a sequence ranging from +1 to -90. This fragement contains the TATA-box consensus (TATATAA), two CAAT-box elements and two *as-1* elements (Fang, et al. Plant Cell 1:141-150 (1989); Benfey, et al. EMBO J.9:1677-1684 (1990); Benfey and Chua, Science 250:959-966 (1990)).

Core promoters have a unique structure and organization at the DNA level. Core promoters in a BDPC may have substantial sequence identity or in one aspect of the invention, be identical. In another aspect, the core promoters of the invention have a sequence homology where promoter sequences have a homology of at least about 30% and include in separate aspects of the invention, at least 5, 10 or 20 bp identical contiguous nucleotides within the core promoter region. In another aspect, the core promoters have a sequence homology where promoter sequences have a homology of at least about 40% and include in separate aspects of the invention, at least 5, 10 or 20 identical contiguous nucleotides within the core promoter region. In another aspect, the core promoters have a sequence homology where promoter sequences have a homology of at least about 50% and include in separate aspects of the invention, at least 5, 10 or 20 identical contiguous nucleotides within the core promoter region.

Studies of protein-DNA interactions indicated that the DNA sequence for a core promoter provides critical binding elements and anchoring points essential for the formation of a productive transcription initiation subcomplex that comprises the RNA polymerase II (RNAPII),

numerous transcription factors (TFIIA, TFIIB, TFIID, CIFs, TAFs) and the TATA-binding protein (TBP) (see review by Zhang, Genome Res. 8:319-326 (1998)). Accordingly, it is easily recognized that a core promoter is one of the prerequisite components in the transcriptional machinery and plays an important role in supporting the precise initiation and synthesis of transcripts.

Sources of core promoters include but are not limited to CaMV 35S, CsVMV, ACT2, PRB1B, octopine synthase promoter, nopaline synthase promoter, manopine synthetase promoter, beta-conglycinin promoter, phaseolin promoter, ADH promoter, heat-shock promoters, developmentally regulated promoters, and tissue specific promoters.

Modified Enhancer Complex

The present invention includes a modified enhancer region, to which two core promoters are fused upstream and downstream thereof to form a BDPC. In another aspect of the invention, the enhancer sequences may have substantial sequence identity or may in one aspect include at least two identical enhancer sequences that are arranged in a tandem orientation. Alternatively, the enhancers of the invention have a sequence homology where enhancer sequences have a homology of at least about 30% and include at least 5 bp identical contiguous nucleotides within the enhancer sequence. More specifically, the 3' end of the first enhancer sequence is linked to the 5' end of the second sequence to form a modified enhancer region in a BDPC.

In yet another aspect of the present invention, each repeated enhancer sequence in a modified enhancer region may correspond to a DNA sequence of about 100 bp to more than about 1.0 kbp in length. The choice for a particular repeat size is preferably based on the desired

transcriptional enhancement and the desired requirements for a specific transgene expression pattern controlled by a particular set of cis-acting elements contained within the enhancer DNA sequence.

In yet another aspect, within a modified enhancer region there may be any number of cis-acting elements that are fully functional to the core promoters used in a BDPC. The cis-acting elements are functional, meaning capable of modulating, including enhancing or down-regulating, the initiation and synthesis of transcripts from a transgene containing either expressible or non-expressible coding sequences.

A modified enhancer region in a BDPC as used herein, may comprise at least two, more than two, or multiple of two, such as four and six, repeated enhancer sequences. If four enhancer repeat sequences are to be used to form a four-unit modified enhancer region in a BDPC, two enhancer sequences are first placed in tandem to form one enhancer array. Two different enhancer arrays made from a total of four repeat sequences will be then fused together in an opposite or back-to-back orientation. More specifically, transcription in the upstream direction may occur on the bottom strand whereas transcription in the downstream direction may occur on the top strand. Likewise, in the case where six enhancer sequences are to be chosen to construct a six-unit modified enhancer region in BDPC, three sequences are first arranged to form an array of tandem repeats. The two different enhancer arrays are finally fused together in a back-to-back orientation to form a six-unit modified enhancer region for use in a BDPC.

The sequence length of all repeated enhancer sequences within one enhancer array may be asymmetrical. As used herein, asymmetrical means that enhancer sequences are at least 10 bp either longer or shorter than the unit length of the enhancer units within the

other enhancer array, as used in either a four- or six-unit modified enhancer region. The use of asymmetric enhancer arrays in a four- or six-unit modified enhancer region is preferred to prevent the formation of a perfect palindromic sequence containing overly long (>100 bp) repeated sequences, which may affect stability during DNA manipulation and cloning processes (Allers and Leach, J. Mol. Biol. 252:72-85 (1995); Nasar et al., Mol. Cell. Biol. 20:3449-3458 (2000)).

The term "enhancer" has been previously defined (Khoury and Gruss, Cell 33:313-314 (1983) and extensively used to describe any DNA sequence with a size ranging from approximately 100 bp to over 2.0 kbp. According to studies of eukaryotic promoters, enhancers are commonly isolated from sequences located upstream or downstream of a core promoter and contain numerous cis-acting elements important for transcription regulation. In an important aspect, enhancers function to modulate, including either enhance or limit, the transcriptional activity of the core promoter in an orientation- and/or position-independent fashion. Transcriptional control or regulation of temporal- and spatial-specific gene expression in all eukaryotes is primarily associated with the presence of functional cis-acting elements within enhancers and is the results of interplay between these regulatory elements and cellular factors in host cells.

Over the years, numerous enhancers have been isolated form organisms ranging from viruses to higher mammals. For instance, in higher plants enhancers regulating gene expression in vegetative tissues, xylem and vascular tissues, roots, flowers, fruits and seeds, as well as gene expression in response to biotic and abiotic stresses, have been isolated and well characterized (see reviews by Edwards and Coruzzi, Annu Rev. Genet. 24:275-303 (1990); Guilfoyle, Genetic Engineering Vol. 19, pps. 15-47 (1997)). Many of these

isolated enhancers have been utilized in efforts to provide regulated control of transgene expression in host and non-host organisms.

Accordingly, in an important aspect of the present invention, all enhancers isolated thus far can be utilized to construct a modified enhancer region for use in a BDPC to effect transgene expression based on the regulatory information contained in the enhancer of choice. Functional enhancers that are chemically synthesized based on predetermined sequence information may also be used in the construction of a modified enhancer region as described in the present invention. The use of repeated enhancers in a modified enhancer region does not alter the gene expression pattern, but primarily provides a unique means to achieve transcriptional enhancement.

DNA can undergo dynamic conformational changes under many circumstances. Certain types of DNA sequences, including tandem repeats, reversed repeats, repetitive sequence arrays, and symmetrical or asymmetrical palindromic sequences, are conducive to the formation of so-called alternative DNA conformations, such as DNA bending, cruciform structures, DNA loops, DNA haripins, DNA 4-way junction structures, DNA triplexes and so forth (Perez et al., Ann. Rev. Microbiol. 51:593-628 (1997); Selker, Cell 97:157-160 (1999); Gaillard et al., BMC Biochem and Struct. Biol. 1:1 (2000); Caddle et al., J. Mol. Biol. 211:19-33 (1990); Courey et al. J. Mol. Biol. 202:35-43 (1988); Spink et al. PNAS 92:10767-10771 (1995); Moore et al. PNAS 96:1504-1509 (1999); Collin et al. NAR 28:3381-3391 (2000)). In some cases, alternative DNA conformations can be derived from intrinsic bonding interactions between nucleic acid residues contained in a unique DNA sequence; in other cases, they may be induced and/or augmented by the interplay between DNA sequence elements and DNA-binding factors (Pil et al. PNAS

90:9465-9 (1993); Wolfe et al. Chem Biol. 2:213-221 (1995); Slama-Schwok et al. NAR 25:2574-81 (1997)). Alternative DNA conformations within eukaryotic enhancers and promoters have been demonstrated to provide important architectural elements, complex signal interaction devices and efficacious molecular environments for DNA-protein interactions that may result in the formation of productive transcriptional machinery (Perz et al. Ann. Rev. Microbiol. 51:593-628 (1997)).

In one aspect, the present invention is intended to introduce into a BDPC an enhancer region modified to contain two tandem repeat(s) of substantially identical enhancer sequences and two core promoters with a high degree of sequence homology placed in opposite orientation on either side of the modified enhancer region. Although any particular helical structure or alternative conformation associated with a BDPC of the present invention needs to be determined by using molecular techniques available in the art, the significant enhancement of transcriptional activity observed from the use of a BDPC suggests the involvement of unique DNA structural geometry that provides a favorable molecular environment for productive interactions between DNA sequence elements within enhancer and core promoters and transcriptional factors present in host cells. Such interactions eventually lead to the onset of synergistically improved transcription from both core promoters.

Transgene Silencing

In another important aspect, the BDPC of the present invention is effective for decreasing the occurrence of gene silencing resulting from loss of promoter function due to methylation and the like. Changes in DNA structure can trigger the onset of gene silencing. Multiple copies of a gene and inverted gene repeats are

vulnerable to DNA methylation modifications that lead to transcriptional silencing (Selker, Cell 97:157-160 (1999)). Tandem repeats of integrated genes can be recognized and modified at the DNA level by host factors (Finnegan et al., Annu. Rev. Plant Physiol. Plant Mol. Biol. 49:223-247 (1998): Kumpatla et al., TIBS 3:97-104 (1998)). A cruciform structure derived from DNA repeats is effectively modified by a mammalian methyltransferase (Smith et al., J. Mol. Biol. 243:143-151 (1994)). However, many cases of transgene silencing derived from repeated sequences involves coding regions (Selker, Cell 97:157-160 (1999); Finnegan et al., Annu. Rev. Plant Physiol. Plant Mol. Biol. 49:223-247 (1998)). BDPCs of the present invention support stable and high levels of transgene expression even though repeated DNA sequences were present within the BDPC region.

Use of BDPCs

In another aspect of the invention, vectors that include a BDPC as described in this invention can be used to express foreign genes in mammalian cells and especially in plant cells that include dicots and monocots. More specifically, dicots include but are not limited to tobacco, grapes, soybeans, legumes, rapeseed, cotton, sunflower, tomatoes, potatoes, sugar beets, alfalfa, cloves and peanuts. Monocots include but are not limited to maize, wheat, sorghum, oats, rye, barley, rice, millets, sugar cane and grasses.

Several techniques exist for introducing foreign genetic material into plant cells, and for obtaining plants that stably maintain and express the introduced gene. Such techniques include acceleration of genetic material coated onto microparticles directly into cells (US Patents 4,945,050 to Cornell and 5,141,131 to DowElanco). Plants may be transformed using Agrobacterium technology, see US Patent 5,177,010 to

University of Toledo, 5,104,310 to Texas A&M, European Patent Application 0131624B1, European Patent Applications 120516, 159418B1, European Patent Applications 120516, 159418B1 and 176,112 to Schilperoot, US Patents 5,149,645, 5,469,976, 5,464,763 and 4,940,838 and 4,693,976 to Schilperoot, European Patent Applications 116718, 290799, 320500 all to MaxPlanck, European Patent Applications 604662 and 627752 to Japan Tobacco, European Patent Applications 0267159, and 0292435 and US Patent 5,231,019 all to Ciba Geigy, US Patents 5,463,174 and 4,762,785 both to Calgene, and US Patents 5,004,863 and 5,159,135 both to Agracetus. Other transformation technology includes whiskers technology, see U.S. Patents 5,302,523 and 5,464,765 both to Zeneca. Electroporation technology has also been used to transform plants, see WO 87/06614 to Boyce Thompson Institute, 5,472,869 and 5,384,253 both to Dekalb, WO9209696 and WO9321335 both to PGS. All of these transformation patents and publications are incorporated by reference. In addition to numerous technologies for transforming plants, the type of tissue which is contacted with the foreign genes may vary as well. Such tissue would include but would not be limited to embryogenic tissue, callus tissue type I and II, hypocotyl, meristem, and the like. Almost all plant tissues may be transformed during dedifferentiation using appropriate techniques within the skill of an artisan.

Foreign genetic material introduced into a plant may include a selectable marker. The preference for a particular marker is at the discretion of the artisan, but any of the following selectable markers may be used along with any other gene not listed herein which could function as a selectable marker. Such selectable markers include but are not limited to aminoglycoside phosphotransferase gene of transposon Tn5 (Aph II) which encodes resistance to the antibiotics kanamycin, neomycin

and G418, as well as those genes which code for resistance or tolerance to glyphosate; hygromycin; methotrexate; phosphinothricin (bar); imidazolinones, sulfonylureas and triazolopyrimidine herbicides, such as chlorosulfuron; bromoxynil, dalapon and the like.

In addition to a selectable marker, it may be desirous to use a reporter gene. In some instances a reporter gene may be used without a selectable marker. Reporter genes are genes which are typically not present or expressed in the recipient organism or tissue. The reporter gene typically encodes for a protein which provide for some phenotypic change or enzymatic property. Examples of such genes are provided in K. Weising et al. Ann. Rev. Genetics, 22, 421 (1988), which is incorporated herein by reference. Preferred reporter genes include without limitation glucuronidase (GUS) gene and GFP genes.

Once introduced into the plant tissue, the expression of the structural gene may be assayed by any means known to the art, and expression may be measured as mRNA transcribed, protein synthesized, or the amount of gene silencing that occurs (see U.S. Patent No. 5,583,021 which is hereby incorporated by reference). Techniques are known for the in vitro culture of plant tissue, and in a number of cases, for regeneration into whole plants (EP Appln No. 88810309.0). Procedures for transferring the introduced expression complex to commercially useful cultivars are known to those skilled in the art.

Once plant cells expressing the gene under control of a bidirectional promoter are obtained, plant tissues and whole plants can be regenerated therefrom using methods and techniques well-known in the art. The regenerated plants are then reproduced by conventional means and the introduced genes can be transferred to

other strains and cultivars by conventional plant breeding techniques.

The following examples illustrate methods for carrying out the invention and should be understood to be illustrative of, but not limiting upon, the scope of the invention which is defined in the appended claims.

EXAMPLES

EXAMPLE 1: Preparation of Transformation Vectors

Two transformation vectors were constructed as illustrated in Fig. 13. Firstly, a green fluorescent protein (GFP) expression cassette was constructed. This cassette was composed of an EGFP (Clontech Laboratories, Inc., Palo Alto, CA) under the control of a core promoter (-90 to +1) (Benfey et al., Science 250:959-966 (1989)), and the terminator and polyadenylation signal of CaMV 35S transcript. This expression cassette was then isolated as a *HindIII* fragment and inserted into the 5' *HindIII* site of the T-DNA region of a binary vector pBI434 (Li et al., Transgenic Crop I. Biotechnology in Agriculture and Forestry, vol. 46 (1999)). This binary vector contained a GUS-NPTII fusion gene (Dalta et al., Gene 101:239-246 (1991)) under the control of an enhanced double CaMV 35S promoter (Kay et al., Science 236:1299-1302 (1987)) followed by a 5' nontranslated leader sequence of alfalfa mosaic virus (AMV) and with a terminator and polyadenylation signal of the nopaline synthase gene of *Agrobacterium*. Two transformation vectors were obtained depending on the orientation of insertion. In vector p201, the GFP expression cassette was in a tandem orientation relative to the GUS-NPTII expression unit. Secondly, the GFP expression cassette in vector p201R was in a divergent orientation leading to the formation of a BDPC in this vector. In the BDPC, two identical core promoters of the CaMV 35S transcript were located on

either side of a duplicated enhancer region [2X (-363 to -91)] resulting in a total size of 736 bp in length (Fig. 2).

EXAMPLE 2: Transformation of Somatic Embryos of Grape

Binary vectors p201 and p201R were both introduced into *A. tumefaciens* strain EHA105 and subsequently used to transform somatic embryos (SE) of grape (*Vitis vinifera* cv. Thompson Seedless). Expression of the EGFP gene was monitored after transformation using a stereomicroscope equipped with a fluorescence illuminator and GFP filter system. GUS expression was quantitatively determined by using a fluorogenic assay as described by Jefferson (Plant Mol. Biol. Rep. 5:387-405).

As shown in Fig. 14, the differential effects of vectors p201 and p201R on the level of GFP expression were readily noticeable one week after transformation. SE transformed with p201 fluoresced only slightly, while SE transformed with p201R fluoresced brightly. Microscopic observation of the SE revealed that the density of GFP-expressing cells on the surface of transformed SE was similar for both vector treatments. These results indicated that the observed difference in the level of GFP expression between these two vectors was the result of the difference in strength of the promoters used to control EGFP gene expression (Fig. 13). The reduced level of GFP expression in SE following transformation with p201, as opposed to p201R, suggests that the transcriptional activity of the same core promoter can be dramatically increased by using a BDPC.

In addition to enhancing gene expression, use of BDPC increased transformation efficiency based on assays of transient GFP expression (Fig. 15). In two independent experiments, transformation using p201R resulted in an increase of about 19% and about 44%,

respectively, in the number of GFP-expressing SE, when compared to p201.

To examine the effect of the BDPC on the downstream core promoter, GFP-expressing SE were selected and further analyzed for GUS expression using a fluorogenic assay. The results illustrated in Fig. 16 indicate that GUS activity in SE transformed using p201R was consistently about 40% higher than the GUS activity detected in SE transformed using p201.

Transgenic embryos and plants were subsequently recovered from the SE transformed using p201R. A consistently high level of GFP expression was observed throughout their subsequent developmental stages and in various plant tissues (Fig. 17), with a similar gene expression pattern achieved by using the CaMV 35S promoter as reported previously (Benfey et al., Science 250:959-966 (1989)). This suggests that the induced enhanced gene expression is spatially and temporally stable in transgenic grape plants.

Experimental data obtained indicate that the BDPC present in p201R is capable of significantly elevating the level of expression of both transgenes (EGFP and GUS), as compared to that obtained using p201, which contains a conventional promoter/transgene configuration. This gene expression enhancement is possibly attributable to an improvement in the structural configuration of the BDPC that results in increased promoter activity.

The addition of a second core promoter to the upstream region of the double promoter in a tandem orientation relative to the downstream core promoter, in p201 constituted an array of tandem repeats of promoter sequences within the T-DNA which induces gene silencing (Kumpatla et al., TIBS 3:97-104 (1998)).

EXAMPLE 3: Quantification of Transgene Expression

To determine quantitatively the transgene expression under control of the upstream core promoter in a BDPC as described in the invention, transformation vectors pLC501T and pLC501R were constructed. As illustrated in Fig. 24, the T-DNA regions of both pLC501T and pLC501R were essentially identical to that of pLC201 and pLC201R, respectively, as shown in Fig. 13, except that the positions of the GUS gene and the EGFP/NPTII gene were switched around, and both transgenes were fused to the terminator of CaMV 35S transcript.

Both pLC501T and pLC501R were introduced into *A. tumefaciens* and subsequently used in transformation of grape SE (cv. Thompson Seedless) as described in Example 2. In this experiment, transformation vector pBI434 containing no BDPC but a GUS/NPTII fusion gene under control of an enhanced double CaMV 35S promoter was also included for GUS activity comparison. Fig. 25 shows GUS activity in SE transformed with various vectors. Noticeably, the core promoter in pLC501T only supported a minimum level of GUS expression (8 pmol MU/mg for 60 min), while a huge increase in GUS expression was observed from SE transformed with pLC501R (1774 pmol MU/mg for 60 min). In other words, up to 220-fold increase in GUS activity was achieved by using pLC501R in which the GUS gene was under the control of the upstream core promoter in a BDPC setting, as compared to the GUS activity derived from the same core promoter without a BDPC configuration (pLC501T). In addition, the GUS activity derived from the upstream core promoter of the BDPC in pLC501R increased by 2-fold, as compared to GUS activity resulted from pBI434, which only contained an enhanced double CaMV 35S promoter. These data, together with observations described in Example 2, clearly demonstrate that a BDPC as described in the invention is effective for achieving stable and significantly high

levels of transgene expression enhancement from both core promoters.

EXAMPLE 4: Quantification of Transgene Expression under 4-Enhancer-Containing BDPC

To investigate transgene expression directed by a BDPC containing 4 enhancers, two transformation vectors pLC903T and pLC903R were constructed. As shown in Fig. 26, both vectors contained an EGFP expression unit and a GUS-containing expression unit. The two expression units were under the control of a similar enhanced double CaMV 35S promoter with a slightly different sequence length of enhancers. In pLC903T the two expression units were placed in a tandem orientation. The two expression units in pLC903R were placed in a divergent (back-to-back) orientation, thus resulting in the formation of a 4-enhancer-containing BDPC for the expression of both EGFP and GUS genes. The BDPC configuration in pLC903R is basically similar to that as illustrated in Fig. 3.

Both pLC903T and pLC903R were introduced into A. *tumefaciens* and subsequently used in transformation of grape SE along with a control transformation vector pBI434 as previously described in Examples 2 and 3. The level of GUS expression in transformed SE was determined subsequently and the averaged results from three independent experiments were summarized in Fig. 27. In these experiments, GUS activity obtained from 30-min reactions was used for data conversion. Results indicated that there was no GUS-specific activity in non-transformed SE (CK-0.3 pmol MU/mg/min). Surprisingly , the GUS activity obtained from SE transformed with pLC903T was about half of that observed from pLC434 (36 vs. 65.4 pmol MU/mg/min), even though the GUS expression unit in both vectors was identical and was controlled by the same enhanced double CaMV 35S promoter. The reduction in GUS expression observed from the use of

pLC903T could be accounted for by the possible interference of terminator sequences (35S-31) in the upstream region of the GUS expression unit in pLC903T. On the contrary, an increase in GUS activity by almost 10-fold was observed in SE transformed with pLC903R, which contains a 4-enhancer-containing BDPC in the upstream region of the core promoter, as compared to the GUS activity from pBI434, which only contained an enhanced double CaMV35S promoter (638.2 vs. 65.4 pmol MU/mg/min). The dramatic increase in GUS expression by using transformation vector pLC903R further demonstrated the significant enhancement of trangene expression from the use of unique BDPC promoter configuration as elucidated in this invention.

Numerous modifications and variations in practice of the invention are expected to occur to those skilled in the art upon consideration of the foregoing detailed description of the invention. Consequently, such modifications and variations are intended to be included within the scope of the following claims.